

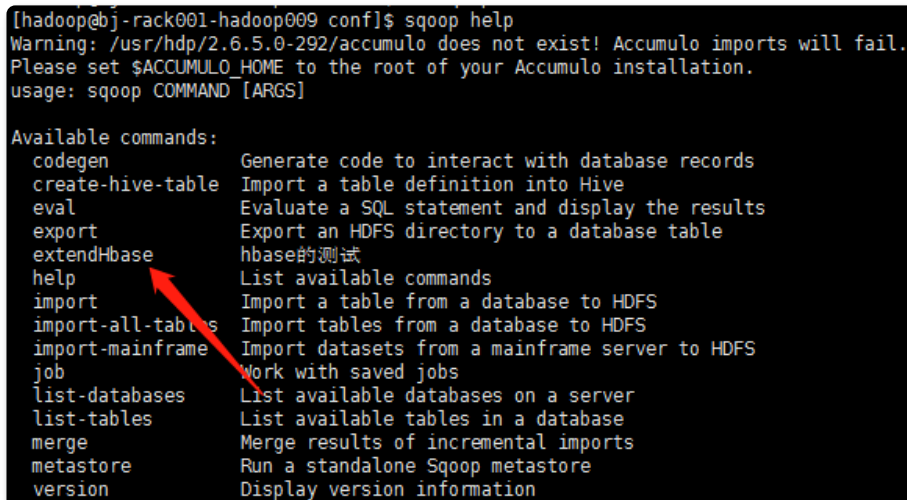
汪帅

注册用户的自定义插件

1. 把自定义的jar文件拷贝到相关集群的 /Sqoop/lib 目录，并在 sqoop-site.xml 进行注册。

```
<property>
  <name>sqoop.tool.plugins</name>
  <value>org.apache.sqoop.extend.ExtendHBasePlugin</value>
  <description>A comma-delimited list of ToolPlugin implementations
    which are consulted, in order, to register SqoopTool instances which
    allow third-party tools to be used.
  </description>
</property>
```

注册结束后再命令行输入：sqoop help 显示自定义的命令即注册成功。



```
[hadoop@bj-rack001-hadoop009 conf]$ sqoop help
Warning: /usr/hdp/2.6.5.0-292/accumulo does not exist! Accumulo imports will fail.
Please set $ACCUMULO_HOME to the root of your Accumulo installation.
usage: sqoop COMMAND [ARGS]

Available commands:
codegen          Generate code to interact with database records
create-hive-table Import a table definition into Hive
eval             Evaluate a SQL statement and display the results
export           Export an HDFS directory to a database table
extendHbase      hbase的测试
help             List available commands
import           Import a table from a database to HDFS
import-all-tables Import tables from a database to HDFS
import-mainframe Import datasets from a mainframe server to HDFS
job              Work with saved jobs
list-databases   List available databases on a server
list-tables      List available tables in a database
merge            Merge results of incremental imports
metastore        Run a standalone Sqoop metastore
version          Display version information
```

Sqoop脚本示例

```
sqoop extendHbase \
  -hbase-table mytable \
  -column-family i \
  -hbase-row-key GDDM,GDXM \
  -export-dir /yss/guzhi/interface/gh.tsv \
  -hbase-rowkey-separator _ \
  -hdfs-line-separator \\t \
  -hbase-col GDDM,GDXM,BCRQ,CJBH,GSMD,CJSL,BCYE,ZQDM,SBSJ,CJSJ,CJJG,CJJE,SQBH,BS,MJBH
```

相关配置的简介

属性	描述
extendHbase	导出HDFS上的数据到HBase的指令
hbase-table	要导入的HBase表名称,需要提前建表不支持自动建表

属性	描述
column-family	HBase表的列簇名称
hbase-row-key	HBase表Rowkey的设置,如果是指定多个列作为RowKey,彼此间的分隔符采用逗号间隔
export-dir	要导出的HDFS上的文件
hbase-rowkey-separator	如果RowKey采用的是复合列组合的方式的默认采用的分隔符是下划线,可以传递参数替换
hdfs-line-separator	HDFS上的文件数据行的分隔符
hbase-col	HBase的列名称与HDFS上的数据列一一对应,彼此之间采用逗号分隔

注意：在设置RowKey时多个字段必须在hbase-col设置中必须包含。